



PATENT ABSTRACTS OF JAPAN

(11) Publication number: **06149482 A**

(43) Date of publication of application: 27.05.94

(51) Int. Cl.

G06F 3/06(21) Application number: **04301367**(22) Date of filing: **11.11.92**(71) Applicant: **HITACHI LTD**

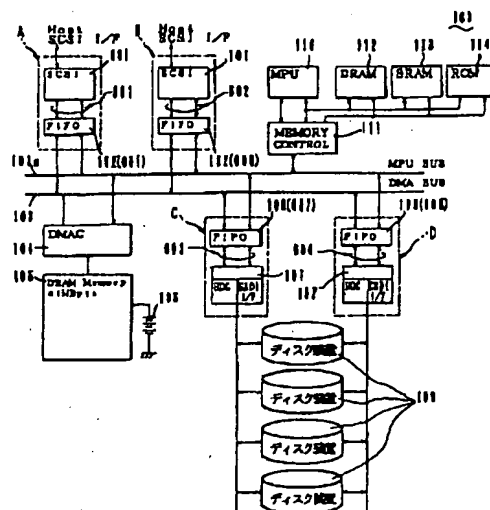
(72) Inventor: **SATO MASAHICO
NAKAKOSHI KAZUO
TAKAHASHI NAOYA
CHUMA AKIRA
YUGAWA YOSHIO**

(54) EXTERNAL STORAGE DEVICE**(57) Abstract:**

PURPOSE: To provide an external storage device capable of shortening time for responding to respective read/write requests from a host device and substantially increasing the issuing number of the read/write requests from the host device.

CONSTITUTION: In a controller part 100 interposed between disk devices 109 and the host device, the buffer memory 105 of a large capacity nonvolatilized by a battery 106, plural host interfaces A and B, plural drive interfaces C and D, a DMA bus 103 and a DMA controller 104 are provided. By the reconnection by dynamic selection and the release of the host interfaces A and B and the drive interfaces C and D, a read/write processing from the host device to a buffer memory 105 and the read/write processing between the buffer memory 105 and the disk devices 109 are parallelly and asynchronously performed.

COPYRIGHT: (C)1994,JPO&Japio



(19)日本国特許庁(JP)

(12) 公開特許公報(A)

(11)特許出願公開番号

特開平6-149482

(43)公開日 平成6年(1994)5月27日

(51)Int.Cl.

G 0 6 F 3/06

識別記号

3 0 2 A 7165-5B

庁内整理番号

F I

技術表示箇所

審査請求 未請求 請求項の数8(全 13 頁)

(21)出願番号 特願平4-301367

(22)出願日 平成4年(1992)11月11日

(71)出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72)発明者 佐藤 雅彦

神奈川県小田原市国府津2880番地 株式会

社日立製作所ストレージシステム事業部内

(72)発明者 中越 和夫

神奈川県小田原市国府津2880番地 株式会

社日立製作所ストレージシステム事業部内

(72)発明者 高橋 直也

神奈川県小田原市国府津2880番地 株式会

社日立製作所ストレージシステム事業部内

(74)代理人 弁理士 筒井 大和

最終頁に続く

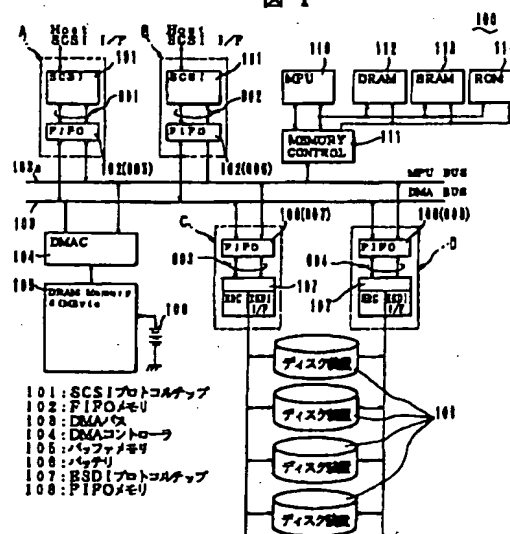
(54)【発明の名称】 外部記憶装置

(57)【要約】

【目的】 上位装置からの個々のリード/ライト要求に対する応答時間の短縮、および上位装置からのリード/ライト要求の発行件数の大幅な増加を実現することが可能な外部記憶装置を提供する。

【構成】 ディスク装置109と上位装置との間に介在するコントローラ部100に、バッテリ106によって不揮発化された大容量のバッファメモリ105と、複数口のホストインターフェイスA、Bおよび複数口のドライブインターフェイスC、Dと、DMAバス103およびDMAコントローラ104を設け、ホストインターフェイスA、BおよびドライブインターフェイスB、Cの解放および動的な選択による再接続により、上位装置からバッファメモリ105に対するリード/ライト処理と、バッファメモリ105とディスク装置109との間のリード/ライト処理を非同期に並行して行うようにした外部記憶装置である。

図 1



【特許請求の範囲】

【請求項1】 データを第1の記憶媒体に持久的に記憶するドライブと、このドライブと上位装置との間における前記データの授受を制御する入出力制御部とを備えた外部記憶装置であって、前記入出力制御部は、前記第1の記憶媒体よりも前記データに対するアクセスをより高速に行うことが可能な第2の記憶媒体からなる一時データ保持手段、および前記上位装置からの前記ドライブに対する前記データの書き込み要求に対し、当該書き込みデータを前記一時データ保持手段に一旦書き込み、当該書き込み動作が完了した時点で前記上位装置に対して書き込み完了を応答する手段、および前記一時データ保持手段から前記ドライブへの前記データの書き込みを、前記上位装置からの前記データの書き込みおよび読み出しのタイミングとは独立に前記入出力制御部で管理されたタイミングで行う手段を含むことを特徴とする外部記憶装置。

【請求項2】 前記入出力制御部は、前記上位装置からの前記データの読み出し要求に対し、前記ドライブから読み出したデータを前記一時データ保持手段に残す手段、および前記上位装置からの同一の前記データの読み出し要求に対し、前記一時データ保持手段に残された前記データを読み出して転送する手段を含むことを特徴とする請求項1記載の外部記憶装置。

【請求項3】 前記第1の記憶媒体が磁気ディスクであり、前記第2の記憶媒体が半導体メモリからなり、当該半導体メモリはセルフ・リフレッシュ機能を有するダイナミック・ランダム・アクセス・メモリ（DRAM）で構成され、前記半導体メモリの電源は、前記入出力制御部と共通の主電源と、この主電源とは独立なバッテリーの2系統で構成され、前記バッテリーの充電が前記主電源によって行われるようにしたことを特徴とする請求項1または2記載の外部記憶装置。

【請求項4】 前記上位装置との間の第1のインターフェイスポートを複数口有し、前記入出力制御部は、前記上位装置から発行される前記データの読み出しまたは書き込みコマンドに対し、当該データの読み出しまたは書き込み準備が完了するまでの間、前記第1のインターフェイスポートを解放する手段、および読み出しまたは書き込みの準備が完了した時点で、前記第1のインターフェイスポートを再接続する手段、および当該再接続に際して、複数の前記第1のインターフェイスポートのうち、空いているポートを選択し、動的に第1のインターフェイスポートの接続を行う手段、および前記動的な第1のインターフェイスポートの接続を抑止する手段を含むことを特徴とする請求項1、2または3記載の外部記憶装置。

【請求項5】 前記入出力制御部に対して前記ドライブをディージーチェーンで複数台接続するとともに、当該入出力制御部と当該ドライブとの間に第2のインターフ

ェイスポートを複数口設け、前記入出力制御部は、この第2のインターフェイスポートによる前記ドライブに対する前記データの読み出しまたは書き込みにおいて、前記ドライブにおける前記データの読み出しまたは書き込み準備が完了するまでの間、当該第2のインターフェイスポートを解放する手段、および前記ドライブにおける前記データの読み出しまたは書き込み準備が完了した時点で、当該ドライブから入力される接続要求割り込みに応じて当該第2のインターフェイスポートを再接続する手段、および当該第2のインターフェイスポートの再接続に際して、複数の前記第2のインターフェイスポートのうち、空いている側を選択して動的に再接続する手段を含むことを特徴とする請求項1、2、3または4記載の外部記憶装置。

【請求項6】 前記入出力制御部は、全体の制御を司るマイクロプロセッサと、複数の前記第1のインターフェイスポートおよび前記第2のインターフェイスポートと前記一時データ保持手段とを時分割処理によって切替えながら接続することによりDMA（ダイレクト・メモリ・アクセス）転送を行うDMAコントローラと、前記マイクロプロセッサと前記第1、第2のインターフェイスポートおよび前記DMAコントローラを接続するMPUバスと、前記DMAコントローラと前記第1、第2のインターフェイスポートを接続するDMAバスと、前記第1および第2のインターフェイスポートと前記MPUバスおよびDMAバスとの間に設けられ、前記一時データ保持手段といずれかのインターフェイスポートとの間でDMA転送が行われている間、他のインターフェイスポートを停止させないためのFIFO（先入れ／先出し）メモリとを含むことを特徴とする請求項1、2、3、4または5記載の外部記憶装置。

【請求項7】 DMAバスのバス幅を、前記第1および第2のインターフェイスポートのバス幅よりも大きくしたことを特徴とする請求項1、2、3、4、5または6記載の外部記憶装置。

【請求項8】 前記DMAバスと、前記第1および第2のインターフェイスポートのバス幅の変換を、前記FIFOメモリの入出力バス幅を当該DMAバス側と当該第1および第2のインターフェイスポート側で変えることにより実現することを特徴とする請求項1、2、3、4、5、6または7記載の外部記憶装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は外部記憶装置に関し、特に、OLTP（Online Transaction Process）、RDB（Relational Data Base）等、高スループット、高速I/Oを要求される情報処理システムに使用される磁気ディスク装置などの外部記憶装置に適用して有効な技術に関する。

【0002】

【従来の技術】高性能のマイクロプロセッサの登場により、計算機システムのデータ処理能力は、急速に高速化され、また、データベース等の大容量化に伴い、これら計算機システムで扱われるデータの規模も大型化の傾向にある。

【0003】このような背景のもと、これらの計算機システムから外部記憶装置の一種である磁気ディスク装置に発行される単位時間当たりのI/O要求件数は、近年急速に大きくなってきている。しかし、磁気ディスク装置は、その動作原理上、データアクセス時に機械的な動作に伴うため、そのアクセス速度は、計算機のCPUな主メモリの速度と比べ大きな隔たりを持っている。このため、磁気ディスク装置が計算機システムの処理能力を決定する上でのボトルネックとなり、計算機システムの性能が磁気ディスク装置のスループットに大きく影響されるといった傾向がある。

【0004】従来、このような問題を改善する手段としてLRU (Least Recently Used)方式を用いたリードキャッシュを用い、記憶媒体上において時間的、物理的近傍でアクセスされたデータはキャッシュメモリに残しておき、上位装置からのリード要求がキャッシュメモリ内のデータであった場合(キャッシュヒット)には、磁気ディスクに対してアクセスすることなくキャッシュメモリ上のデータを転送し、応答時間の短縮を図ることが行われていた。また、特開昭60-7457号公報に開示される技術では、複数台の磁気ディスク装置に複数台の磁気ディスクコントローラを接続し、デバイスクロスコール方式を改善した技術が提示されているが、この技術も同様にLRU方式を用いたリードキャッシュ方式の応用にしかすぎず、データのライト処理に対しては磁気ディスク装置の性能の低さがそのままシステムの処理能力に反映されてしまう、といった不具合があった。

【0005】また、従来の磁気ディスクコントローラは、上位装置からのリードおよびライトコマンドに対し、下位ディスク装置にシークコマンドを発行し、ヘッドが目的のトラックに到達するまでのシーク時間、および目的のトラック上のセクタがヘッド下に到来するまでの回転待ち時間の間、インターフェイスを接続状態のままとしており、上位装置から極めて高頻度のI/O要求が発行された場合、インターフェイスのバスネックとなり、I/O発行件数が制約されるという問題があった。

【0006】

【発明が解決しようとする課題】磁気ディスク装置に対する高速化の要求として、I/O性能の向上、すなわち、高速にI/O処理を実行すること、および、単位時間当たりのI/O発行件数の向上がある。

【0007】上記従来技術では、いずれの場合もデータの読み出し処理についてはそれなりの効果があるものの、データの書き込みについては配慮されておらず、磁気ディスク装置のトータルのI/O性能としては、必ず

しも計算機システムの要求性能を満足するものではなかった。

【0008】また、近年、計算機システムの外部記憶装置に要求される記憶容量の増大、およびクライアント/サーバ方式などに代表される分散処理の流れから、磁気ディスクサブシステムに複数台の磁気ディスクドライブを内蔵することで、大容量化を実現し、このような構成のサブシステムに対して、複数台のホストコンピュータからのアクセス要求が発行される様な環境下で磁気ディスクサブシステムが稼動することが現実となってきている。

【0009】このような状況から、磁気ディスクサブシステム内のコントローラは、複数台のホストコンピュータと、複数台の磁気ディスクドライブの接続を制御しなくてはならなくなって来ている。

【0010】上記従来技術では、ホストインターフェイスおよびドライブインターフェイスは、それぞれディーゼーチェーン方式で複数台のホストコンピュータおよび複数台の磁気ディスクドライブに接続され、特定のホストコンピュータからのI/O要求が発生した場合には、当該I/O要求が終了するまで、それぞれのホストコンピュータは待ち状態となり、インターフェイスのバスネックによりI/O発行件数が制限されるといった問題があった。この対策として、たとえばインターフェイスバスを個別に持たせることも考えられるが、コントローラが非常に高価かつ大規模になってしまうという他の問題を生じてしまう。

【0011】本発明の目的は、上位装置からの個々のリード/ライト要求に対する応答時間の短縮、および上位装置からのリード/ライト要求の発行件数の大幅な増加を実現することが可能な外部記憶装置を提供することにある。

【0012】本発明の前記ならびにその他の目的と新規な特徴は、本明細書の記述および添付図面から明らかになるであろう。

【0013】

【課題を解決するための手段】本願において開示される発明のうち、代表的なものの概要を簡単に説明すれば、下記のとおりである。

【0014】すなわち、本発明は、データを第1の記憶媒体に持久的に記憶するドライブと、このドライブと上位装置との間におけるデータの授受を制御する入出力制御部とを備えた外部記憶装置において、入出力制御部は、第1の記憶媒体よりもデータに対するアクセスをより高速に行うことが可能な第2の記憶媒体からなる一時データ保持手段、および上位装置からのドライブに対するデータの書き込み要求に対し、当該書き込みデータを一時データ保持手段に一旦書き込み、当該書き込み動作が完了した時点で上位装置に対して書き込み完了を応答する手段、および一時データ保持手段からドライブへの

データの書き込みを、上位装置からのデータの書き込みおよび読み出しのタイミングとは独立に入出力制御部で管理されたタイミングで行う手段、を有したものである。

【0015】また、本発明は、請求項1記載の外部記憶装置において、入出力制御部には、上位装置からのデータの読み出し要求に対し、ドライブから読み出したデータを一時データ保持手段に残す手段、および上位装置からの同一のデータの読み出し要求に対し、一時データ保持手段に残されたデータを読み出して転送する手段を設けたものである。

【0016】また、本発明は、請求項1または2記載の外部記憶装置において、第1の記憶媒体が磁気ディスクであり、第2の記憶媒体が半導体メモリからなり、当該半導体メモリはセルフ・リフレッシュ機能を有するダイナミック・ランダム・アクセス・メモリ(DRAM)で構成され、半導体メモリの電源は、入出力制御部と共通の主電源と、この主電源とは独立なバッテリーの2系統で構成され、バッテリーの充電が主電源によって行われるようにしたものである。

【0017】また、本発明は、請求項1、2または3記載の外部記憶装置において、上位装置との間の第1のインターフェイスポートを複数口有し、入出力制御部には、上位装置から発行されるデータの読み出しまたは書き込みコマンドに対し、当該データの読み出しまたは書き込み準備が完了するまでの間、第1のインターフェイスポートを解放する手段、および読み出しまたは書き込みの準備が完了した時点で、第1のインターフェイスポートを再接続する手段、および当該再接続に際して、複数の第1のインターフェイスポートのうち、空いている

ポートを選択し、動的に第1のインターフェイスポートの接続を行う手段、および動的な第1のインターフェイスポートの接続を抑止する手段を設けたものである。

【0018】また、本発明は、請求項1、2、3または4記載の外部記憶装置において、入出力制御部に対してドライブをディージーチェーンで複数台接続するとともに、当該入出力制御部と当該ドライブとの間に第2のインターフェイスポートを複数口設け、入出力制御部には、この第2のインターフェイスポートによるドライブに対するデータの読み出しまたは書き込みにおいて、ドライブにおけるデータの読み出しまたは書き込み準備が完了するまでの間、当該第2のインターフェイスポートを解放する手段、およびドライブにおけるデータの読み出しまたは書き込み準備が完了した時点で、当該ドライブから入力される接続要求割り込みに応じて当該第2のインターフェイスポートを再接続する手段、および当該第2のインターフェイスポートの再接続に際して、複数の第2のインターフェイスポートのうち、空いている側を選択して動的に再接続する手段を備えたものである。

【0019】また、本発明は、請求項1、2、3、4ま

たは5記載の外部記憶装置において、入出力制御部は、全体の制御を司るマイクロプロセッサと、複数の第1のインターフェイスポートおよび第2のインターフェイスポートと一時データ保持手段とを時分割処理によって切替えながら接続することによりDMA(ダイレクト・メモリ・アクセス)転送を行うDMAコントローラと、マイクロプロセッサと第1、第2のインターフェイスポートおよびDMAコントローラを接続するMPUバスと、DMAコントローラと第1、第2のインターフェイスポートとを接続するDMAバスと、第1および第2のインターフェイスポートとMPUバスおよびDMAバスとの間に設けられ、一時データ保持手段といずれかのインターフェイスポートとの間でDMA転送が行われている間、他のインターフェイスポートを停止させないためのFIFO(先入れ/先出し)メモリとを含む構成としたものである。

【0020】また、本発明は、請求項1、2、3、4、5または6記載の外部記憶装置において、DMAバスのバス幅を、第1および第2のインターフェイスポートのバス幅よりも大きくしたものである。

【0021】また、本発明は、請求項1、2、3、4、5、6または7記載の外部記憶装置において、DMAバスと、第1および第2のインターフェイスポートのバス幅の変換を、FIFOメモリの入出力バス幅を当該DMAバス側と当該第1および第2のインターフェイスポート側で変えることにより実現するものである。

【0022】

【作用】通常、たとえば磁気ディスク装置などの外部記憶装置においては、データの書き込みは、ヘッドを目的のシリンダに移動し、さらにディスクの回転によってヘッドが目的のセクタ上に到達した後に行われる。この場合、ヘッドを移動するシーク時間、およびヘッドの直下に目的のセクタが到来するまでの回転待ち時間などの機械的な動作時間がオーバーヘッドとして加算され、書き込み時のI/O性能を制限している。

【0023】上記した本発明の外部記憶装置では、ドライブの第1の記憶媒体よりも遙に高速な半導体メモリなどからなる一時データ保持手段に対して、リードデータのみならず、書き込み要求に際してはライトデータも格納し、当該一時データ保持手段に対するライトデータの書き込みが完了した時点で、上位装置に対して書き込み動作の完了を報告し、一時データ保持手段からドライブへの実際の書き込みは、上位装置からの書き込み要求の契機とは非同期に実行するので、ドライブ側における前述のような機械的な要因の各種待ち時間が書き込みデータの処理の所要時間に影響することがなくなり、上位装置から発行される個々のI/O要求をドライブの動作性能に無関係に高速に処理することができる。

【0024】また、下位のドライブとの間の第2のインターフェイスポートにおいて、ドライブに対してコマン

ドを発行した時点で当該ポートを解放し、ドライブ側の準備が完了した時点で再接続の割り込み要求を発行して当該ポートを再接続する動作を行うので、あるドライブの準備の間は第2のインターフェイスポートを他のドライブとのI/O処理に用いることが可能となる。また、さらに、第2のインターフェイスポートを複数備え、前記再接続などに際しては空いているインターフェイスポートを動的に選択して接続することにより、複数の第2のインターフェイスポートを無駄なく使用でき、しかも再接続の待ち時間も短縮される。この結果、単位時間当たり10に受付処理可能なI/Oの発行件数が増加し、複数の上位装置によって共有される場合に発生するデバイスクロスコールなどに的確に対応することができる。

【0025】

【実施例】以下、本発明の一実施例である外部記憶装置を図面を参照しながら詳細に説明する。

【0026】図1は本実施例の外部記憶装置の構成の一例を示すブロック図である。本実施例では、外部記憶装置の一例として、磁気ディスクサブシステムの場合について説明する。

【0027】本実施例の磁気ディスクサブシステムは、たとえば磁気ディスクなどを記憶媒体としてデータを持久的に記憶する複数台のディスク装置109と、このディスク装置109と図示しないホストコンピュータ（上位装置）（以下、ホストと略記する）との間におけるデータの授受を制御するコントローラ部100とを含んでいる。

【0028】コントローラ部100のMPUバス103 aおよびDMAバス103には、図示しないホストとの間に介設される2系統のホストインターフェイスA、ホストインターフェイスB、および複数台のディスク装置109がディジーチェーン接続される2系統のドライブインターフェイスCおよびドライブインターフェイスDが接続されている。

【0029】複数のホストインターフェイスA、Bの各々は、ホストとの間でSCSI (Small Computer System Interface) プロトコルによる情報の授受を制御するSCSIプロトコルチップ101と、このSCSIプロトコルチップ101とMPUバス103 aおよびDMAバス103との間に介在するFIFOメモリ102によって構成されている。

【0030】一方、複数のドライブインターフェイスC、Dの各々は、コントローラ部100に接続される複数のディスク装置109との間において、ESDI (Enhanced Small Device Interface) プロトコルによって情報の授受を行うESDIプロトコルチップ107と、このESDIプロトコルチップ107とMPUバス103 aおよびDMAバス103との間に介在するFIFOメモリ108によって構成されている。

【0031】この場合、MPUバス103 aおよびDM 50

Aバス103には、DMAコントローラ104を介してバッファメモリ105が接続されている。このバッファメモリ105は、たとえば容量が64メガバイトの半導体メモリ（セルフ・リフレッシュ機能付のDRAM）で構成されている。さらに、バッファメモリ105は図示しない主電源と、当該主電源によって常時充電されているバッテリー106の双方から給電される構成となっており、主電源が切断された場合、バッテリー106からの給電によって、たとえば最長一週間、記憶データを保持することが可能になっている。

【0032】また、MPUバス103 aには、バスコントローラ111を介して、本実施例の磁気ディスクサブシステムの全体の制御を行うマイクロプロセッサ110 (MPU)、当該マイクロプロセッサ110の制御プログラムやワーク用メモリエリアを提供する制御メモリ112 (DRAM)、制御メモリ113 (SRAM)、制御メモリ114 (ROM) が接続されており、後述のような一連の制御動作が実現される構成となっている。

【0033】以下、本実施例の磁気ディスクサブシステムの作用の一例について説明する。

【0034】図2および図3に、図示しない上位のホストからディスク装置109に対してデータを書き込む場合のデータの流れを示す。

【0035】図2は、ホストからバッファメモリ105にデータを転送するまでの処理を示している。ホストからデータの書き込み要求が発生した場合、SCSIプロトコルチップ101から割り込み信号201が発生し、マイクロプロセッサ110はデータの書き込み要求を認識する。マイクロプロセッサ110は、この要求に従い、バスコントローラ111およびMPUバス103 aを介して制御アクセス202、制御アクセス203、制御アクセス204を行うことにより、SCSIプロトコルチップ101、FIFOメモリ102、DMAコントローラ104に対してDMA転送のための各種レジスタの設定などを行う。

【0036】以上の設定が終了すると、ホストからのデータは、バス206でバッファメモリ105に対してDMA転送される。また、このDMA転送の間、マイクロプロセッサ110は、制御アクセス205により、ESDIプロトコルチップ107を介してディスク装置109に対して、データを格納すべきシリンダへのシーク命令の発行およびヘッドの切替えなどを行う。

【0037】ホストからのDMA転送と、ディスク装置109におけるシークが終了した時点でマイクロプロセッサ110は、次に、バッファメモリ105に格納された書き込みデータ207をディスク装置109に転送する。

【0038】この時の書き込みデータ207の流れを図3に示す。マイクロプロセッサ110は、制御アクセス301、制御アクセス302、制御アクセス303によ

り、DMAコントローラ104、ESDIプロトコルチップ107、FIFOメモリ108に対して、DMA転送のための設定を行う。以上の設定が終了すると、バッファメモリ105上の書き込みデータ207は、バス304でバッファメモリ105からディスク装置109に転送され、目的のディスク装置109の格納位置305に書き込まれる。

【0039】以上の制御を図4のタイムチャートで示す。同図では、SCSIプロトコルチップ101、DMAバス103、ディスク装置109(Drive)の各々における処理動作を並べて示している。書き込み要求が発生すると、ホストは、コントローラ部100に対して区間401の間にデータ転送を行う。これと同時に、コントローラ部100の内部でSCSIプロトコルチップ101からバッファメモリ105に対してDMA転送が開始され、区間402でデータ転送が行われる。DMA転送が若干遅れるのは、FIFOメモリ102でデータをバッファリングしているためで、この遅れ時間は、転送所要時間に比べて極めて短い時間である。また、前記転送動作と同時にディスク装置109に対してシークコマンドが発行され、ディスク装置109は、区間403でシークを実行し、シークが完了するとさらに区間404で回転待ちを行い、データを書き込むべきセクタがヘッドの直下に到来すると区間405でバッファメモリ105からディスク装置109にDMA転送が開始され、区間406でディスク装置109に対するデータの書き込みが実行される。

【0040】従来のディスクコントローラでは、ホストからデータを受理し、ディスク装置109に対するデータの書き込みが完了した区間407でコマンド終了報告を応答していたため、転送開始から区間407までの区間408が一回の書き込み動作の所要時間となっていた。

【0041】これに対して、本実施例の場合には、区間402のDMA転送が終了し、バッファメモリ105に対する書き込みデータの転送完了の区間409でホストに対して書き込みコマンドの終了報告を行っている。このため、ホストから見れば、ディスク装置109におけるシークの区間403と回転待ちの区間404を省いた区間410が一回の書き込みコマンドの所要時間となり、従来に比較して、極めて高速にコマンドが実行されたことになる。ディスク装置109への書き込みは、ホストに対する応答処理とは非同期に実行される。すなわち、ホストからのI/O要求が混み合っている場合は、ディスク装置109へのデータの書き込みは一時保留し、I/O要求頻度が低下した時点で実行することでさらにホストからのコマンドを高速に処理することができる。

【0042】バッファメモリ105に空き領域が無くなった場合は、未反映のデータのディスク装置109への

書き込み終了を待ってから、ホストからのデータの受領を行うため、従来と同程度のスループットに低下するが、本実施例のように、バッファメモリ105の容量を大きく取った場合、当該事象は極めて多くのI/O要求が発行されるまで発生することはない。

【0043】ここで問題となることは、ホストからのバッファメモリ105への書き込みデータのDMA転送が終了し、ホストにコマンド終了報告を行った後、データがディスク装置109に実際に書き込まれる前に、なんらかの原因で主電源が切断された場合のデータ喪失の危険性である。

【0044】本実施例の場合、前述のようにバッファメモリ105をバッテリー106によってバックアップすることで、不揮発性メモリとし、主電源の切断時においても、バッファメモリ105の内容が約一週間保持されるため、データ喪失に至ることはない。ただし、信頼性の確保のために、通常の稼働において主電源を切断する際は、バッファメモリ105上のデータをすべてディスク装置109に反映させた後、主電源を切断する手続が採られる。

【0045】ディスク装置109からホストにデータを読み出す場合は、前述のデータ書き込みの処理とは逆の処理が行われる。すなわち、ディスク装置109からバッファメモリ105上にデータを読み出し、当該バッファメモリ105に目的のデータがセーブされた時点でバッファメモリ105からホストにデータが転送される。この場合は、ディスク装置109に対するアクセスが先となるため、最初の読み出し処理では書き込みの場合のような早期応答が行えないが、たとえば、同一データに対する2回目以降の読み出しで、仮に、バッファメモリ105にデータが存在した場合、すなわちキャッシュヒットの場合には、ディスク装置109からの実際の読み出し動作を行うことなく、高速にアクセス可能なバッファメモリ105のデータを転送できるため、やはり、極めて高速なコマンドの実行が可能となる。

【0046】すなわち、本実施例によれば、ホストからのI/O要求と、実際のディスク装置109に対するアクセスが分離され、ホストからのI/O要求は実質的に高速なバッファメモリ105との間で行われるので、ホストとの間で高スループットのデータ転送を行うことができる。

【0047】次に、データ書き込み時にバッファメモリ105に空き領域が無くなった場合、およびデータの読み出し時にバッファメモリ105でミスヒットが発生し、ホストからのアクセス要求にディスク装置109のアクセスが伴う場合を考える。このような事象が発生するケースは、一般に複数のホストが磁気ディスクサブシステムを共有するマルチホスト構成で、極めて高頻度のI/O要求が発行される場合と考えられる。

【0048】このような条件下では、単にディスク装置

10

20

30

40

50

109のシーク時間や回転待ち時間がコマンドの実行時間に含まれることによるコマンドの実行速度の低下の他に、多発するI/O要求の衝突によるホストインターフェイスA、BおよびドライブインターフェイスC、D（インターフェイスバス）の空き待ちによるスループットの低下が発生する。すなわち、バスネックによるスループットの低下および単位時間当たりのI/O処理数の低下である。この問題を解決するために、本実施例では、複数のドライブインターフェイスC、Dを設けるとともに、コントローラ部100に、ディスク装置109がシーク中および回転待ち時にドライブインターフェイスC、Dを解放する制御論理、およびデータ転送時に任意の空いている側のドライブインターフェイスCまたはDを用いて動的に再接続する制御論理を持たせ、バスネックによるスループットの低下を回避している。

【0049】図5に、このような制御論理による処理のタイムチャートの一例を示す。

【0050】同図の例では、4台のホスト（Host 0、1、2、3）と4台のディスク装置109（Drive 0、1、2、3）との間で、Host 0がDrive 0に、Host 1がDrive 1に、Host 2がDrive 2に、Host 3がDrive 3にリードコマンドを発行し、すべてミスヒットとなり、ディスク装置109に対するアクセスが必要となった場合を想定している。

【0051】まず、区間501でHost 0からリードコマンドが発行されると、コントローラ部100は区間502で一方のホストインターフェイスA（SCSI 0）からコマンドを受領し、区間503でDrive 0に対し、ドライブインターフェイスC（ESDI 0）を介してシークコマンドを発行する。Drive 0は、区間504でシークを実行する。このシーク中はSCSI 0およびESDI 0は使われないため、区間505、区間506で当該ESDI 0およびSCSI 0は解放される。

【0052】引続き、区間507でHost 1からリードコマンドが発行されると、コントローラ部100は、区間508で、空いているSCSI 0から当該コマンドを受領し、Drive 1に対し区間509で空いているESDI 0を介してシークコマンドを発行する。Drive 1は区間510でシークを実行する。そして、Drive 0の場合と同様に、区間511、区間512でESDI 0およびSCSI 0は解放される。

【0053】次に、Host 1によるリードコマンドの発行の区間507に僅かに遅れた区間513でHost 2からリードコマンドが発行される。

【0054】このとき、SCSI 0およびESDI 0はHost 1からのコマンド発行に使用中であるため、ホストインターフェイスB（SCSI 1）からコマンドが発行され、コントローラ部100は区間514でコマン

ドを受理する。そして、目的のDrive 2に対して区間515でドライブインターフェイスD（ESDI 1）を介してシークコマンドを発行する。Drive 2は、区間516でシークを実行し、さらに、区間517、区間518でESDI 1およびSCSI 1は解放される。

【0055】次に、区間519でHost 3からリードコマンドが発行される。このとき、SCSI 0、ESDI 0はすでに解放されているため、空いているSCSI 0から当該リードコマンドは発行され、コントローラ部100は区間520で当該リードコマンドを受理する。そして、Drive 3に対して区間521でESDI 0を介してシークコマンドを発行する。Drive 3は区間522でシークを実行する。このシーク中、区間523および区間524でESDI 0およびSCSI 0は解放される。

【0056】コマンド発行が終了後、各Drive 0～3のシークが終了し、ヘッドが目的のセクタに達すると、各Drive 0～3からコントローラ部100に対して割り込みが発行され、これを契機に、コントローラ部100は、Drive 0～3およびHost 0～3の再接続を行い、データの転送を行う。

【0057】図5の例では、最初にDrive 0のシークが終了した場合を示している。

【0058】区間525はDrive 0の回転待ちを示し、区間526でDrive 0からデータ転送が行われる。このデータは、区間527、区間528でESDI 0およびSCSI 0を再接続し、区間529でHost 0に送られる。

【0059】引続き、Drive 1がシーク終了し、区間530での回転待ち時間後、当該Drive 1は区間531でデータ転送を開始する。この時、ESDI 0およびSCSI 0はDrive 0からのデータ転送で占有されているため、空いている側のESDI 1、SCSI 1が区間532、区間533で再接続され、区間534でHost 1に送られる。

【0060】Host 1からのリードコマンドは、SCSI 0、ESDI 0をから発行されていたが、再接続時には、他の空いている側のSCSI 1およびESDI 1を選択することにより、従来のような固定的なインターフェイスの設定によるデータ転送におけるバスネックが解消される。

【0061】なお、本実施例のコントローラ部100では、このホストインターフェイスA、BおよびドライブインターフェイスC、Dの動的な切り離しおよび再接続を可能または不能に設定する制御論理を持っており、必要に応じて、データ転送中に使用する各インターフェイスを固定することも可能である。

【0062】以下、同様に、Drive 2、Drive 3からの再接続を行い、全てのデータ転送が終了する。

【0063】従来のように、Host 側およびDrive

10

20

30

40

50

e側に単一のインターフェイスしかなく、しかも当該各インターフェイスの動的な切り離しや再接続の機能がな
い場合、上述したHost 0~3とDrive 0~3との間におけるデータ転送処理は全てシーケンシャルとなり、本実施例の場合に比較して、約2.5~3倍の処理時間
を要することとなる。

【0064】ここで、図5の区間535に注目すると、この間、SCSI 0~1およびESDI 0~1のすべてのインターフェイスにおいてDMA転送が行われるため、バッファメモリ105に対するアクセスが同時に進行することになる。

【0065】そこで、これを同時処理するため、本実施例では、ホストインターフェイスA (SCSI 0)、ホストインターフェイスB (SCSI 1) およびドライブインターフェイスC (ESDI 0)、ドライブインターフェイスD (ESDI 1) の各々にFIFOメモリ102およびFIFOメモリ108を設け、DMA転送はサイクルスチールで時分割処理を行う。

【0066】図6にDMA転送の経路周辺に着目した概念図を示す。

【0067】FIFOメモリ605 (102)、FIFOメモリ606 (102) の各バス601およびバス602は、それぞれSCSI 0、SCSI 1に繋がり、FIFOメモリ607 (108)、FIFOメモリ608 (108) の各バス603およびバス604は、ESDI 0、ESDI 1に繋がる。

【0068】DMAコントローラ104は、FIFOメモリ605~608のうちのいずれか一つを選択し、DMAバス103を介してバッファメモリ105との間でデータ転送制御を行う。バス601~604による転送を連続同時に行うためには、DMAバス103のデータ転送速度はバス601~604の全てのバス転送速度の合計を上回る必要がある。

【0069】本実施例のコントローラ部100では、Host側に接続されるバス601、602は幅が16ビットで最大10MB/secであり、Drive側に接続されるバス603、604は幅が8ビットで最大5MB/secであり、DMAバス103は、30MB/sec以上の転送速度が必要となる。このような高速転送を実現するため、本実施例のコントローラ部100では、DMAバス103の幅を32ビットとしており、当該DMAバス103とバス601、602およびバス603、604の幅変換を、FIFOメモリ605~608で実現している。

【0070】以上説明したように、本実施例の外部記憶装置によれば、ホスト側から見たアクセス要求に対する応答時間の大幅な短縮、さらには単位時間当たりのI/O要求発行件数を大幅に向上させることができるという効果が得られる。

【0071】たとえば、本発明者らのシミュレーション

によれば、本実施例で開示した構成の磁気ディスクサブシステムでは、バッファメモリを持たず、また単一のホストインターフェイスおよびドライブインターフェイスを持つ構成とした従来技術に比較して、平均発行I/O件数で60%の向上、平均応答時間で約57%の低減が期待できることが確認されている。

【0072】以上本発明者によってなされた発明を実施例に基づき具体的に説明したが、本発明は前記実施例に限定されるものではなく、その要旨を逸脱しない範囲で種々変更可能であることはいうまでもない。

【0073】

【発明の効果】本願において開示される発明のうち、代表的なものによって得られる効果を簡単に説明すれば、以下のとおりである。

【0074】すなわち、本発明の外部記憶装置によれば、上位装置からの個々のリード/ライト要求に対する応答時間の短縮、および上位装置からのリード/ライト要求の発行件数の大幅な増加を実現することができるという効果が得られる。

20 【図面の簡単な説明】

【図1】本発明の一実施例である外部記憶装置の構成の一例を示すブロック図である。

【図2】本発明の一実施例である外部記憶装置の作用の一例を示す概念図である。

【図3】本発明の一実施例である外部記憶装置の作用の一例を示す概念図である。

【図4】本発明の一実施例である外部記憶装置の作用の一例を示すタイムチャートである。

30 【図5】本発明の一実施例である外部記憶装置の作用の一例を示すタイムチャートである。

【図6】本発明の一実施例である外部記憶装置において、DMA転送の経路に着目した概念図である。

【符号の説明】

100 コントローラ部 (入出力制御部)

101 SCSIプロトコルチップ

102 FIFOメモリ (605~606)

103 DMAバス

103a MPUバス

104 DMAコントローラ

40 105 バッファメモリ (一時データ保持手段) (第2の記憶媒体)

106 バッテリ

107 ESDIプロトコルチップ

108 FIFOメモリ (607~608)

109 ディスク装置 (Drive 0~3) (第1の記憶媒体)

110 マイクロプロセッサ

111 バスコントローラ

112 制御メモリ (DRAM)

113 制御メモリ (SRAM)

(9)

特開平 6-149482

15

16

114 制御メモリ (ROM)
 201 割り込み信号
 202~205 マイクロプロセッサ110の制御アクセス
 206 バス
 207 書き込みデータ
 301~303 マイクロプロセッサ110の制御アクセス
 304 バス
 305 ディスク装置でのデータの格納位置
 401 ホストからの書き込みデータ転送
 402 SCSIプロトコルチップからバッファメモリへのDMA転送
 403 ディスク装置 (Drive) でのシーク処理
 404 回転待ち時間
 405 バッファメモリからディスク装置 (Drive) へのDMA転送
 406 ディスク装置へのデータ転送
 407 ホストへのコマンド終了報告 (従来技術の場合)
 408 コマンド実行時間 (従来技術の場合)
 409 ホストへのコマンド終了報告 (本発明の場合)
 410 コマンド実行時間 (本発明の場合)
 501 Host 0からDrive 0へのリードコマンド発行
 502 501のコマンド発行に対するSCSI 0の占有区間
 503 501のコマンド発行に対するESDI 0の占有区間
 504 501のコマンド発行に対するDrive 0のシーク
 505 501のコマンド発行に対するESDI 0の解放
 506 501のコマンド発行に対するSCSI 0の解放
 507 Host 1からDrive 1へのリードコマンド発行
 508 507のコマンド発行に対するSCSI 0の占有区間
 509 507のコマンド発行に対するESDI 0の占有区間
 510 507のコマンド発行に対するDrive 1のシーク
 511 507のコマンド発行に対するESDI 0の解放
 512 507のコマンド発行に対するSCSI 0の解放
 513 Host 2からDrive 2へのリードコマンド発行

10

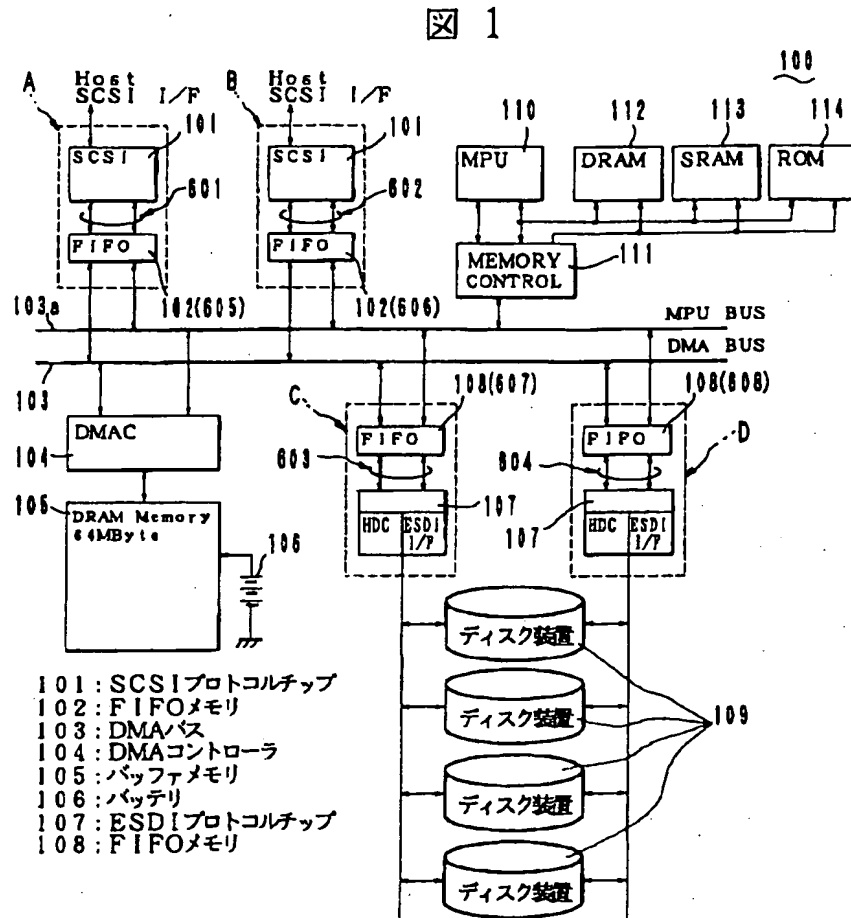
20

30

40

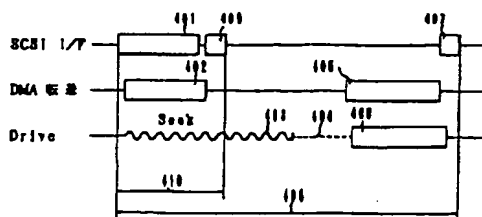
514 513のコマンド発行に対するSCSI 1の占有区間
 515 513のコマンド発行に対するESDI 1の占有区間
 516 513のコマンド発行に対するDrive 2のシーク
 517 513のコマンド発行に対するESDI 1の解放
 518 513のコマンド発行に対するSCSI 1の解放
 519 Host 3からDrive 3へのリードコマンド発行
 520 519のコマンド発行に対するSCSI 0の占有区間
 521 519のコマンド発行に対するESDI 0の占有区間
 522 519のコマンド発行に対するDrive 3のシーク
 523 519のコマンド発行に対するESDI 0の解放
 524 519のコマンド発行に対するSCSI 0の解放
 525 Drive 0のシーク後の回転待ち時間
 526 Drive 0からのデータ読み出し
 527 526のデータ読み出しに対するESDI 0上でのデータ転送
 528 526のデータ読み出しに対するSCSI 0上でのデータ転送
 529 526のデータ読み出しに対するHost 0のデータ受理
 530 Drive 1のシーク後の回転待ち時間
 531 Drive 1からのデータ読み出し
 532 531のデータ読み出しに対するESDI 1上でのデータ転送
 533 531のデータ読み出しに対するSCSI 1上でのデータ転送
 534 531のデータ読み出しに対するHost 1のデータ受理
 535 SCSI 0~1およびESDI 0~1のすべてでDMA転送が行われる区間
 601~604 FIFOメモリのバス
 A ホストインターフェイス (SCSI 0) (第1のインターフェイスポート)
 B ホストインターフェイス (SCSI 1) (第1のインターフェイスポート)
 C ドライブインターフェイス (ESDI 0) (第2のインターフェイスポート)
 D ドライブインターフェイス (ESDI 1) (第2のインターフェイスポート)

【図1】

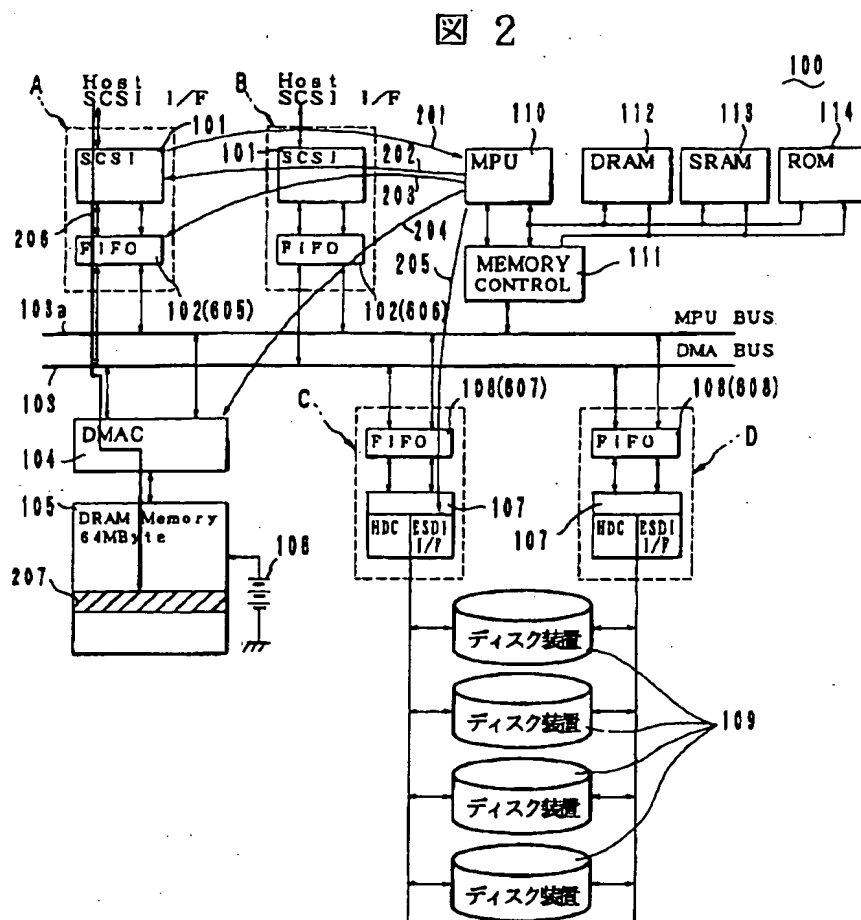


【図4】

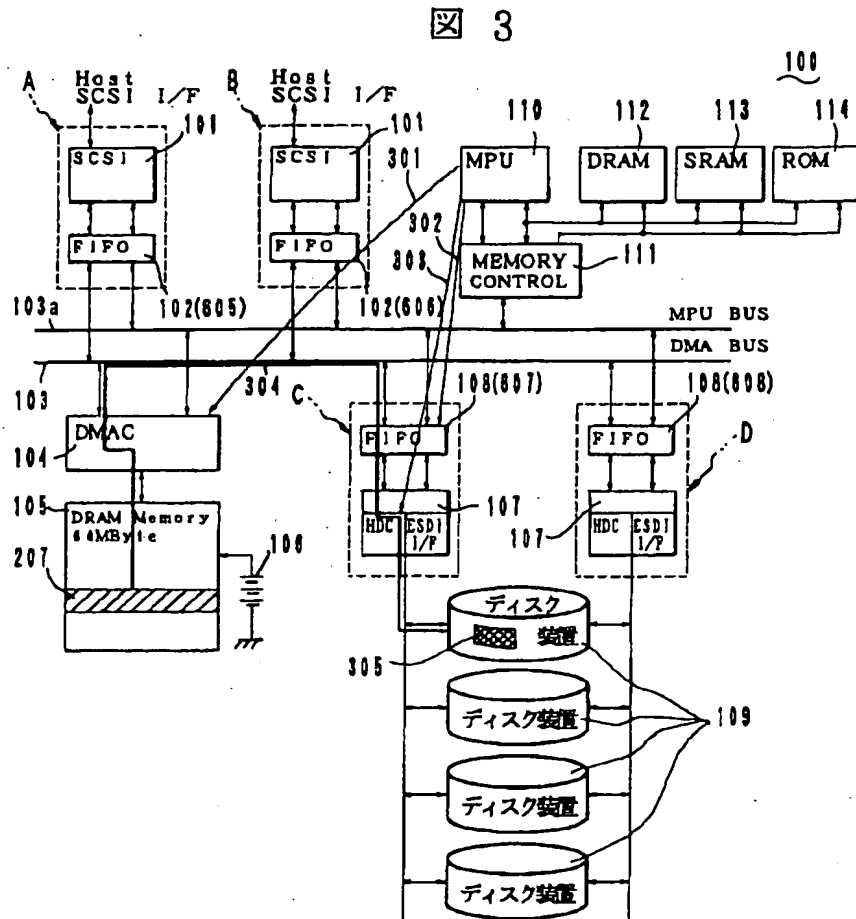
図 4



【図 2】



【図3】



【図5】

【図6】

図 5

